DOCUMENT RESUME

ED 022 533 PS 001 123

By-Kelley, Marjorie L.; Chen, Martin K.

AN EXPERIMENTAL STUDY OF FORMAL READING INSTRUCTION AT THE KINDERGARTEN LEVEL.

Spons Agency-Rosenburg Foundation, San Francisco, Calif.
Pub Date [62]
Note-6p.
EDRS Price MF-$0.25 HC-$0.32
Descriptors-ANALYSIS OF VARIANCE, ATTITUDE TESTS, GROUP INTELLIGENCE TESTING, *KINDERGARTEN CHILDREN, RATING SCALES, *READING INSTRUCTION, *READING READINESS, READING READINESS TESTS, READING SKILLS, READING TESTS, *SCHOOL ATTITUDES, TEACHER ATTITUDES
Identifiers-California Reading Test, California Test Of Mental Maturity, Lee Clark Reading Readiness Test

For 4 months, 221 kindergarten children took part in a controlled experiment on the effects of teaching formal reading at the kindergarten level. Teachers involved in the study were judged comparable in efficiency and attitudes towards the children. The California Test of Mental Maturity Pre-primary Kindergarten 1 was administered to half of the children and the Lee Clark Reading Readiness Test was given to the other half. High IQ, low IQ, high readiness, and low readiness groups emerged. Children were randomly assigned to formal reading or readiness programs. At the end of the school year the children were given the California Reading Test and the School Attitude Inventory. The teachers filled out rating scales on each child's attitude towards school. Results of analysis of variance of the data imply that the two types of instruction operated uniformly on all four categories of subjects. Children in the formal reading program surpassed children in the readiness instruction program in reading skills. According to teachers' ratings, attitudes towards school and reading were a function of intelligence and reading readiness. However, on the pupil self-reporting scale, children in the readiness group had more favorable attitudes towards school. (MS)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OE EDUCATION POSITION OR POLICY.

PAPER NO. 222

RECEIVED

FEB 5 1968

An Experimental Study of Formal Reading Instruction at the Kindergarten Level[1]

Marjorie L. Kelley and Dr. Martin K. Chen
Livermore School District, California

ED022533

ED022533

PS001123

ABSTRACT

The purpose of the study was to investigate the effects of formal reading instruction on kindergartners with respect to reading achievement, attitude toward reading, and attitude toward school.

Altogether 220 children, classified on intelligence and reading readiness variables, were randomly assigned to formal reading and readiness programs for four months. Criterion data respecting to achievement and attitudes were collected by means of the California Reading Test and constructed attitude inventories.

Analyses of data reveal that in terms of reading achievement, the reading program was more effective than the readiness program, but that attitudes toward school and reading were a function of intelligence and reading readiness when attitudes were measured by a teacher-reporting scale. When measured by a pupil self-reporting scale, attitudes were a function of the type of instruction, with children in the readiness program showing more favorable attitudes.

IN THE LAST few years, there has been increasing interest in not only "how" to teach children to read, but "when" to teach them to read. Both popular magazines and professional journals are carrying a proliferation of articles on opinions, hypothesis and research on when to teach reading.

Some parents and some professional educators have been forging ahead with reading instruction, beginning at four or five years of age, rather than waiting for the traditional age of six years and six months.

One national survey indicates a teaching trend toward earlier reading for some children. The efficacy of teaching reading earlier is still in question (1).

While over 3,000 studies in reading have been made, few have been done with the stringent experimental controls that are necessary to give validity to the findings. This investigation has been undertaken with a view to exploring various questions of formal reading instruction at the kindergarten level and to providing some answers, however tentative, to some of the most common questions that are generally asked by educators and by interested parents of kindergarten children.

Statement of the Problem

The problem in which the investigators were primarily interested was whether or not formal reading instruction at the kindergarten level would increase the children's reading skills and favorably dispose them toward reading activities and the school. And if it did, would it benefit the children of different intelligence and readiness levels equally or differentially? If differentially, just what category or categories of kindergartners benefit the most by formal reading instruction? To answer questions such as these the research design was formulated and implemented.

Procedure

In the fall of 1963, some 221 children enrolled in the kindergarten classes of two schools in the Livermore School District in California were used as subjects in a controlled experiment on the teaching of formal reading at the kindergarten level. The two schools were chosen largely because the teachers from the schools involved in the project were judged comparable in teaching efficiency and in their attitudes toward the children. This attempt to control variation among the teachers was made to ensure internal validity of the study —somewhat at the expense of external validity. That is, it was judged more important to reduce the size of the experiment error that might mask real experimental differences than to ensure that the findings be generalized to other schools in the system. To achieve the latter goal, selection of a random sample of schools would have been necessary (2).

As a further means of error reduction in the experiment, one half of Ss (randomly selected) were given the California Test of Mental Maturity, Form 1957S, Pre-primary Kindergarten 1, and the other half were given the Lee–Clark Reading Readiness Test. On the basis of the test scores, Ss were divided into "high" and "low" categories on intelligence and reading readiness, using the median score as the point of division. This resulted in four categories of Ss: the

"high-IQ" group, the "low-IQ" group, the "high-readiness" group, and the "low-readiness" group. Then, group by group, Ss were randomly assigned to the two treatments of formal reading instruction and the customary kindergarten readiness instruction. In the readiness program the children were made "ready" to read with the aid of picture books, records, etc., but were not formally taught reading. Schematically, the research design appears as follows:

Category of Ss		B_1	B_2	B_3	B_4
Teaching Method	A_1	27	22	26	24
	A_2	25	23	26	24

Where B_1 = high-IQ, B_2 = low-IQ, B_3 = high-readiness and B_4 = lowreadiness, and A_1 = formal reading instruction, and A_2 = readiness instruction.

The unequal cell frequencies resulted from the invalidation of the test scores of some Ss.

Just before the end of the spring semester, 1964, Ss were given the California Reading Test, Form X, the Gates Primary Reading Test, Form 3, the Check List of Reading Habits,[2] the Teacher Rating Scale on Children's Attitude Toward School,[3] and the School Attitude Inventory.[4] Scores derived from the Gates Test were discarded because the test proved too difficult for Ss (too many zero scores), making the test invalid for this specific investigation.

Statistical Analyses

Since the cell frequencies were unequal in the 2 x 4 analysis of variance schema, and since the unequal frequencies resulted from random invalidation of the test score of some Ss, the unweighted means method, rather than the least squares solution, was used in the statistical analyses (3). Individual comparisons were made on a selected basis after visual inspection of the cell means. These posteriori tests were made, not to find significant results, but to isolate the optimal category by treatment combinations. The within-cell mean square that was used as error estimate in the analysis of variance was also used in the t test in individual comparisons.

The analysis of variance indicates that the reading achievement of the kindergartners in the formal reading program, averaged over category, is significantly higher than that of the subjects in the readiness instruction program. Further, the analysis reveals that Ss in different categories achieved differentially. The nature of the differential achievement is brought out in the individual comparisons, by which it is shown that the "high-IQ" categories outperformed the "low-IQ" category in both the formal reading program

Table 1.—Analysis of Variance When California Reading Test Scores Were Used as the Dependent Variable

Source of Variation	SS	df	MS	F
A (method of instruction)	.197	1	.197	6.16 $p < .001$
B (category of Ss)	2.068	3	.689	21.53 $p < .001$
A × B	0.09	3	.030	0.94 NS
Within	6.17	189	.032	

and the readiness instruction program, and the "high-readiness" groups bettered the "low-readiness" groups in both programs. There is no significant difference between the "high-IQ" category, and the "high-readiness" category in either the formal reading program or the readiness instruction program, nor is there any significant difference between the "low-IQ" category and the "low-readiness" category in either program. The assumption is here made that the California Test of Mental Maturity and the Lee–Clark Readiness Test to a large extent tap the same skills of abilities.

The lack of A × B interaction clearly demonstrates that neither intelligence nor readiness as measured by the tests interacts with the type of instructional program. In other words, insofar as the California Reading Test reflects true reading achievement, formal reading instruction benefits all Ss except the low intelligence group, which seemed to come out better under readiness instruction than its counterpart in the formal reading program. This reversal, however, could have been a matter of chance, since the difference is not statistically significant.

Both the F tests and the t tests indicate that the children's reading habits are entirely a function of their intelligence or readiness level. Those in the high intelligence or readiness category have better reading habits than those in the low intelligence or readiness category. The type of

Table 2.—Selected Individual Comparisons When California Reading Test Scores Were Used as the Dependent Variable

Individual Comparisons	t	df	p
A_1B_1 vs. A_2B_1	1.85	189	NS
A_1B_3 vs. A_1B_4	4.44	189	<.01
A_2B_1 vs. A_2B_2	3.16	189	<.01
A_1B_4 vs. A_2B_4	1.32	189	NS
A_1B_1 vs. A_1B_2	5.18	189	<.01
A_1B_3 vs. A_2B_3	2.25	189	<.05
A_2B_3 vs. A_2B_4	3.60	189	<.01
A_1B_2 vs. A_2B_2	−0.23	189	NS

Table 3.—Analysis of Variance When the Check List of Reading Habits Scores Were Used as the Criterion

Source of Variation	SS	df	MS	F
A (method of instruction)	.246	1	.246	NS
B (category of Ss)	3867.51	3	1289.17	25.61 p <.001
A × B	371.44	3	127.14	2.53 NS
Within	9512.13	189	50.32	

Table 4.—Selected Individual Comparisons When the Check List of Reading Habits Scores Were Used as the Criterion

Comparisons	t	df	p
A_1B_1 vs. A_2B_1	1.96	189	NS
A_2B_4 vs. A_1B_4	1.16	189	NS
A_2B_3 vs. A_1B_3	1.478	189	NS
A_1B_1 vs. A_1B_2	5.29	189	<.01
A_1B_3 vs. A_1B_4	3.82	189	<.01
A_2B_1 vs. A_2B_2	2.92	189	<.01
A_2B_3 vs. A_2B_4	4.07	189	<.01

instruction apparently has no bearing on their reading habits. Again the assumption is borne out that the California Test of Mental Maturity and the Lee–Clark Readiness Test tap the same skills or abilities.

Again the analysis of variance and the t tests for mean differences indicate that whether or not the children have a favorable attitude toward school is a function of their intelligence or readiness level. Children in the "high" categories have better attitudes toward school than children in the "low" categories, regardless of the type of instruction they had. However, while the difference between the high and low readiness groups in the reading program is significant at the .01 level, that between the two groups in the readiness instruction program is significant at the .10 level, not the .05 level. This phenomenon seems to be a reflection of the tendency for the readiness

Table 5.—Analysis of Variance When Scores on Teacher Rating Scale on Children's Attitude Toward School Were Used as the Criterion

Source of Variation	SS	df	MS	F
A (method of instruction)	4.43	1	4.43	NS
B (category of Ss)	740.42	3	246.80	12.60 p <.001
A × B	46.52	3	12.14	NS
Within	3685.30	189	19.49	

Table 6.—Selected Individual Comparisons When Scores on the Teacher Rating Scale on Children's Attitude Toward School Were Used as the Criterion

Comparison	t	df	p
A_1B_2 vs. A_2B_2	.81	189	NS
A_1B_4 vs. A_2B_4	1.17	189	NS
A_1B_1 vs. A_1B_2	3.54	179	<.01
A_2B_1 vs. A_2B_2	3.18	189	<.01
A_1B_3 vs. A_1B_4	1.69	189	NS
A_2B_3 vs. A_2B_4	3.52	189	<.01

factor to be inoperative among children in the readiness program, for unspecifiable reasons.

When scores on the School Attitude Inventory were used as the dependent variable, only the A mean square was significant, indicating that the children under the two instructional programs, averaged over category, differed significantly. Visual inspection of the cell means reveals that kindergartners of all categories in the readiness instruction program came out better than those in the formal reading program, though selected individual comparisons indicate that only in the high intelligence category is the difference statistically significant. Of further interest is the fact that neither intelligence nor readiness as a factor differentiates the children in either of the two instructional programs.

Discussion of Findings

One interesting feature of the statistical analyses is that in none of the F tests is the A × B interaction mean square significant. That this is the case not only proves the correctness of the use of the additive model in the analysis of vari-

Table 7.—Analysis of Variance When the School Attitude Inventory Scores Were Used as the Criterion

Source of Variation	SS	df	MS	F
A (method of instruction)	43.32	1	43.32	5.65 p <.01
B (category of Ss)	2.95	3	0.98	NS
A × B	8.86	3	2.33	NS
Within	1447.58	189	7.66	

Table 8.—Selected Comparisons When School Attitude Inventory Scores Were Used as the Criterion

Comparison	t	df	p
A_2B_1 vs. A_1B_1	2.07	189	<.05
A_2B_4 vs. A_1B_4	1.31	189	NS

ance for the scale used, but also points up the lack of significant interaction between category and type of instruction. Geometrically, this lack of interaction is reflected by the expression of the main effects of A or B as more or less parallel lines. In some cases the lines may cross, but since the interaction is not significant, the profiles should look similar. For example, for the data used in Tablt 1, the profile of the A and B main effects looks as follows:

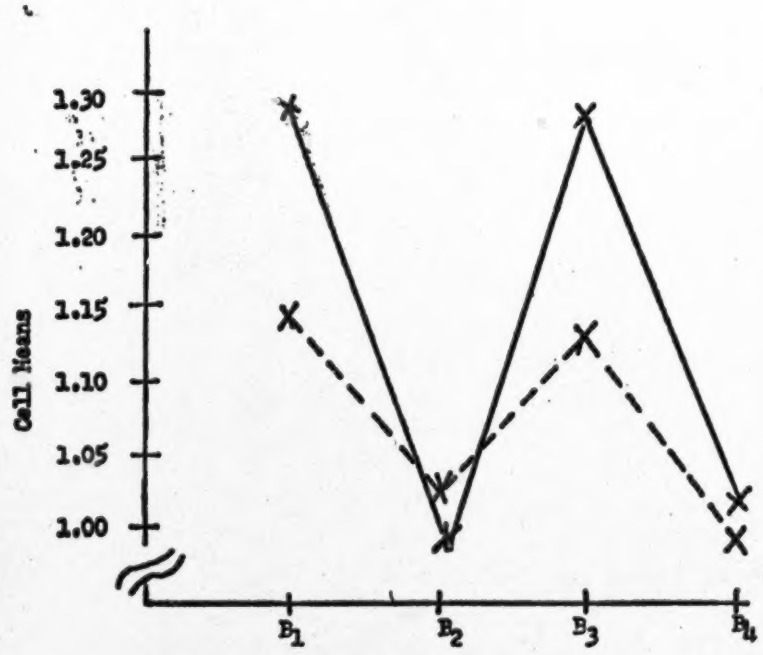

Figure 1.—Profile of A Main Effects

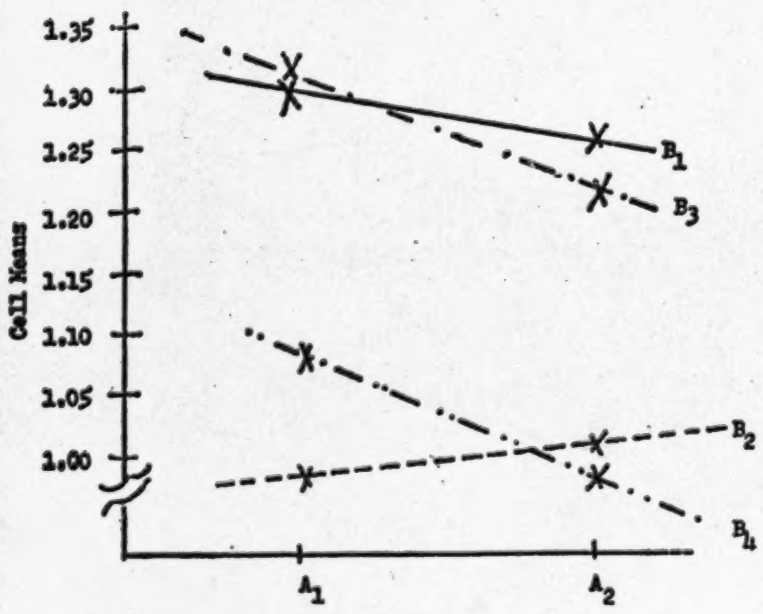

Figure 2.—Profile of B Main Effects

That in both profiles the lines crossed show some interaction effect, though this effect is not statistically significant. In other words, any interaction effect is merely a matter of chance. That is true of all four analyses of variance. The implication of this finding is that the two types of instruction —formal reading or readiness instruction, operate uniformly on all four categories of Ss. For example, Table 7 shows that all four types of Ss seemed to suffer from the formal reading program so far as school attitude was concerned, but Table 8 indicates that the high intelligence category seemed to be most severely affected.

Another interesting finding is that the California Test of Mental Maturity and the Lee–Clark Readiness Test seem to measure the same skills or abilities. A factor analysis would probably bear this out, but even without this analysis it seemed safe to conclude that whichever factor or factors these two tests measure are highly loaded in these two instruments. The implication of this is that in future studies only one of the tests should be used, since the two tests practically duplicate each other's results.

A puzzling phenomenon is the finding that whereas the children's attitudes toward school are a function of their intelligence or readiness level when measured by the Teacher Rating Scale, they become a function of the type of instruction when measured by the School Attitude Inventory. Assuming that the instruments are sufficiently reliable and valid, what factor or factors account for this phenomenon? Obviously, these two instruments do not measure the same attitude, or if they do, they do it on two different dimensions. In the case of the Teacher Rating Scale, the children's scores depend on teacher observation, whereas in the case of the inventory, their scores reflect the degree of the accuracy of self-reporting. A few sample items from both instruments will help to clarify this point.

From Teacher Rating Scale:

1. When the child comes to school, he or she is:

0	.5	1	.5	2	.5	3
Apathetic		Sometimes enthusiastic		Usually enthusiastic		Intensively enthusiastic

4. Toward the teacher, the child appears:

0	.5	1	.5	2	.5	3
Hostile		Indifferent		Fairly friendly		Very friendly

6. When the child is instructed to do something, he or she:

0	.5	1	.5	2	.5	3
Shows strong resistance		Does it reluctantly		Works with indifference		Works with enthusiasm

From the inventory:

2. School is a good place. Yes ( ) Not sure ( ) No ( )
10. There are many interesting things to do in school. Yes ( ) Not sure ( ) No ( )
14. I like talking with the teacher. Yes ( ) Not sure ( ) No ( )

These and other similar items indicate that the theoretical bases of the two instruments are two entirely different cognitive structures—one adult and the other childish. Interestingly, the results of the F tests for Table 3, where scores on the Check List of Reading Habits were the criterion, and for Table 5, where scores from Teacher Rating Scale were the criterion, were identical. That this is so may be explained by the fact that the two instruments were predicated on the same theoretical basis of adult cognitive structure. In both cases the teachers rated the children according to what they observed to be the children's reaction in a certain situation.

On the other hand, the School Attitude Inventory scores are based on what the children feel toward certain objects or activities. Since at this age level (five to six years) they are unlikely to differentiate clearly, at least in terms of the printed word, what is socially acceptable and what is not, they probably cannot fake very well, and their scores can be considered free of variation contributed by faking.

To go back to Tables 3 and 5, it is reasonable to expect that the brighter children would exhibit behaviors which the teachers would interpret as being fond of reading activities and school. At least this is what the analyses of variance seem to bear out. As to why the children in the reading program seemed to suffer in terms of the School Attitude Inventory scores (Table 7), no concrete reasons can be offered that are completely satisfactory. Nevertheless, it may be assumed that the children in the formal reading program were subjected to a tighter schedule than those in the readiness instruction program, with the result that the former had less time for the "free" activities. This might have made school less attractive to them than it was to the children in the readiness instruction program. Other studies have certainly indicated similar results.

PS 001123

Summary and Suggestions for Future Research

Some 221 kindergartners registered in two schools in the Livermore School District were used as subjects in a controlled experiment on the teaching of formal reading at the kindergarten level. Ss were divided into four categories on the basis of their high or low standing on the California Test of Mental Maturity, Form 1957S and the Lee–Clark Readiness Test. Then category by category Ss were randomly assigned to the two treatments of readiness instruction and formal reading instruction.

At the close of the experiment, it was found that Ss in the formal reading program excelled Ss in the readiness instruction program in reading skills. In terms of reading habits, or attitudes toward reading, it was found that they were entirely a function of their intelligence or readiness levels, the type of instruction having no bearing on them. This was also true of their attitudes toward school when they were measured by a teacher rating scale. When their attitudes were assessed by a self-reporting inventory, however, the finding was that Ss in the readiness instruction program had more favorable attitudes toward school than Ss in the formal reading program, irrespective of intelligence or readiness levels. The findings support the assumption that the California Test of Mental Maturity and the Lee–Clark Readiness Test largely tap the same skills or abilities.

While in this study every effort was made to reduce experimental error, no attempt was made to control variability in the home background influence on the subjects. This omission was made for two reasons. First, to obtain data on the children's home background would have taken too much time to get the project underway on schedule. Secondly, it was hoped that the randomization procedure used in assigning Ss to treatments would sufficiently reduce this random error to keep the power of the F tests. Nevertheless, since at this age level the home exercises much more influence on the children than does the school, effective control over this variable would undoubtedly increase the sensitivity of the statistical analyses. It is therefore suggested that in any future study the factor of home background be taken into consideration in the research design. Perhaps a twofold table classification, in which the subjects are divided into four categories of high-IQ and high home background, high-home background, and low-IQ and low home back-home background, and low-IQ and low home background, would be a desirable way of making use of the home background information. Or, more preferably, both intelligence and home background can be used as classification variables in a three-factor factorial design in which treatments constitute levels of the third factor. In this way some information on the two and three-factor interactions would be a bonus over and above the desired information on treatments.

FOOTNOTES

1. This study was made possible with the financial assistance of the Rosenberg Foundation. The authors acknowledge the help of Dr. M. J. Homfeld, Superintendent of the Livermore Elementary School District, Dr. T. B. Edwards and Dr. D. H. Russell of the University of California, Berkeley.

2. Reprinted with permission of the author, Dr. Daniel Johnson, from his unpublished thesis, "Reading Behavior, Achievements and Attitudes of First Grade Boys," Stanford University, February, 1959. No information on validity or reliability is given, but the check list correlates in the high 80's with a reading attitude inventory constructed for the experiment.
3. This is an instrument constructed on a modified principle of the Thurstone method. It has an odd-even reliability of .78. It is not validated for lack of appropriate external criteria.
4. This instrument is of the Likert type. It has an odd-even reliability of .84, but no data on validity.

REFERENCES

1. Hildreth, G. *Readiness for School Beginners* (New York: World Book Co., 1950).
2. Lindquist, E. F. *Design and Analysis of Experiments in Psychology and Education* (New York: Houghton Mifflin Co., 1953).
3. Winer, B. J. *Statistical Principles in Research Design* (New York: McGraw–Hill Book Co., 1962).